

A product rule for Quantum Metrology

Giulio Chiribella

Center for Quantum Information, Institute for Interdisciplinary Information Sciences,
Tsinghua University, Beijing 100084, China.

Abstract. We investigate the optimal estimation of a quantum process that can possibly consist of multiple time steps. The estimation is implemented by a quantum network that interacts with the process by sending an input and processing the output at each time step. We formulate the search of the optimal network as a semidefinite program and use duality theory to give an alternative expression for the maximum payoff achieved by estimation. Combining this formulation with a technique devised by Mittal and Szegedy we prove a general product rule for the joint estimation of independent processes, stating that the optimal joint estimation can be achieved by estimating each process independently, whenever the figure of merit is of a product form. We illustrate the result in several examples and exhibit counterexamples showing that the optimal joint network may not be the product of the optimal individual networks if the processes are not independent or if the figure of merit is not of the product form. In particular, we show that entanglement can reduce by a factor K the variance in the estimation of the sum of K independent phase shifts.

1. Introduction

Quantum theory offers impressive advantages over classical theory in the estimation of physical parameters [1, 2, 3, 4, 5, 6, 7, 10, 11, 12, 13]. The prototypical example is the estimation of an unknown phase shift [3, 4, 11, 12]: here the variance vanishes as N^{-2} with the number N of accesses to the phase-shifting process, whereas a classical statistics over independent copies would give the scaling N^{-1} . The quadratic improvement is achieved by preparing an entangled state of N systems and applying the unknown process to each system. The same quadratic advantage can be found in the estimation of a direction in space [5, 6] and in the joint estimation of three Cartesian axes [7, 8, 9].

Given the usefulness of entanglement in the estimation of a single parameter from multiple accesses to a physical process, it is natural to ask whether entanglement can improve the estimation of many parameters corresponding to different processes. For example, one may wonder whether entanglement can help in the estimation of two independent phase shifts. In a slightly different context, this type of question was originally addressed by Wootters in an unpublished work and by DiVincenzo, Terhal, and Leung [14], who asked whether a joint entangled measurement can improve the extraction of information about two bits encoded in two independent sets of states. In this scenario, it was shown that the amount of information that can be extracted from the product set is additive [14]. More recently, a different proof showing the optimality of product measurements for the extraction of information from general product sets of states was provided in Ref. [15].

In this paper we address the problem of the joint estimation of the parameters encoded in a set of independent processes, where each process can consist of several time steps. Due to the possibility of connecting an input of an unknown process with the output of another one, here the question whether quantum correlations can improve the estimation is not only a question about the usefulness of entanglement in the input states and in the measurements, but also a question about the usefulness of quantum correlations *in time*, namely correlations mediated by the exchange of quantum systems from one time step to the next. We address the question in the framework of quantum estimation [16, 17], where the figure of merit is the expected payoff associated to a payoff function $g(\hat{x}, x)$, which depends of the true value x and of estimated value \hat{x} labelling the unknown process. In this context we prove a general product theorem, showing that the optimal joint estimation of a set of independent parameters $\mathbf{x} := (x_1, \dots, x_K)$ can be achieved by estimating each parameter independently whenever the figure of merit is of the product form $g(\hat{\mathbf{x}}, \mathbf{x}) = \prod_{k=1}^K g_k(\hat{x}_k, x_k)$, where g_k is the payoff function for the parameter x_k . In particular, our result implies that the maximum probability of success in identifying a set of unknown processes is the product of the maximum probabilities of success in identifying each individual process separately.

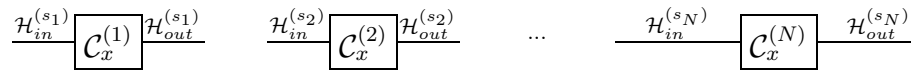
Product theorems are a key tool in theoretical computer science [18, 19, 20, 21, 22, 23, 24], where one is often interested in how the resources needed to solve several independent problems jointly are related to the resources needed to solve each problem

individually. Our work begins to explore the usefulness of this techniques in the domain of physics, starting from the fundamental problem of identifying a set of independent physical parameters. In order to prove our result we use the framework of *quantum combs* [25, 26] (see also the work by Gutoski and Watrous on *quantum strategies* [27]). In this framework we formulate the maximization of the expected payoff as a semidefinite program, and present an intuitive formulation of the dual minimization program. Such a dual formulation is interesting in its own right, as it generalizes to arbitrary processes and arbitrary payoff functions a classic formula derived by Yuen, Kennedy, and Lax [28] for the minimum error state discrimination. Exploiting the form of the primal and dual programs, we then prove our product theorem following a general technique devised by Mittal and Szegedy in Ref. [23] (see also Ref. [24]), which is adapted here in order to deal with the optimization of quantum networks consisting of multiple time steps.

2. Quantum networks for process estimation

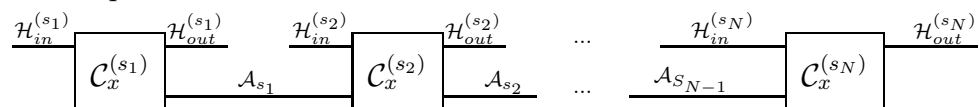
Suppose that an experimenter has access to a physical process \mathcal{P}_x that depends on an unknown parameter x in some parameter space \mathbf{X} . The goal of the experimenter is to determine the parameter x with the maximum precision allowed by the laws of quantum mechanics.

Generally, the process \mathcal{P}_x can consist of N time steps, labelled by an index s in some finite set $\mathbf{S} = (s_1, \dots, s_N) \subset \mathbb{N}$, ordered so that $s_m < s_n$ for $m < n$. At each time step $s \in \mathbf{S}$ the process transforms an input quantum system, with Hilbert space denoted by $\mathcal{H}_{in}^{(s)}$, into a (possibly different) output quantum system, with Hilbert space denoted by $\mathcal{H}_{out}^{(s)}$. If the process \mathcal{P}_x is memoryless, all time steps are independent and one can associate a quantum channel to each time step. The quantum channel at step s , denoted by $\mathcal{C}_x^{(s)}$, will be a completely positive trace-preserving map sending density matrices on $\mathcal{H}_{in}^{(s)}$ to density matrices on $\mathcal{H}_{out}^{(s)}$. Hence, the process \mathcal{P}_x can be described by a time-ordered sequence of quantum channels, each channel labelled by the unknown parameter x , as in the following picture:



In the easiest case, one may have the same channel at each time step, namely $\mathcal{C}_x^{(s)} = \mathcal{C}_x$ for every $s \in \mathbf{S}$. This is the case, e.g. of quantum phase estimation [3, 4, 10, 11, 12, 13], where one has access to N uses of the unitary channel $\mathcal{C}_x = U_x \rho U_x^\dagger$, with $U_x = \exp(ixH)$ for some Hamiltonian H with integer spectrum.

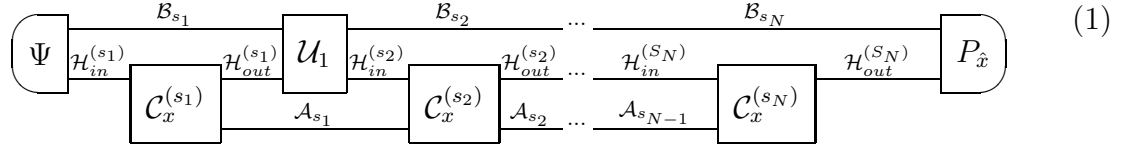
In the presence of memory, the input-output transformation at the step s is described by a quantum channel involving internal ancillas: in this case the quantum channel $\mathcal{C}_x^{(s)}$ transforms density matrices on $\mathcal{H}_{in}^{(s)} \otimes \mathcal{A}_{s-1}$ to density matrices on $\mathcal{H}_{out}^{(s)} \otimes \mathcal{A}_s$, where \mathcal{A}_s is the Hilbert space of the s -th ancilla. Hence, the process \mathcal{C}_x is represented by a time-ordered sequence of black boxes with internal memories:



Note that, since the ancillas are internal to the network, the first and last ancillary systems are trivial $\mathcal{A}_0 \simeq \mathcal{A}_N \simeq \mathbb{C}$.

The most general strategy to estimate an unknown parameter from a time-ordered sequence of black boxes consists in inserting the black boxes in a quantum network where the black boxes are interspersed with known quantum gates and eventually a quantum measurement is performed on the output, producing the estimate $\hat{x} \in \mathbb{X}$.

The estimation process can be depicted as



where \mathcal{B}_s , $s \in \mathcal{S}$ are the internal ancillas of the estimating network, Ψ is a quantum state on $\mathcal{B}_{s_1} \otimes \mathcal{H}_{in}^{(s_1)}$, each \mathcal{U}_s is a quantum channel, and $P_{\hat{x}}$ is a quantum measurement, described by a *positive operator valued measure (POVM)* on the Hilbert space $\mathcal{B}_{s_N} \otimes \mathcal{H}_{out}^{(s_N)}$.

Examples of quantum networks for the estimation of unknown parameters can be found in Refs. [11, 12].

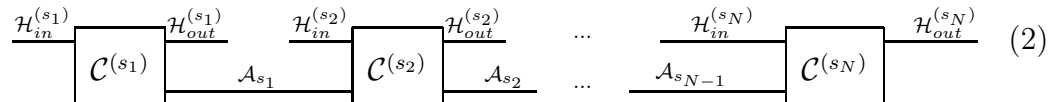
3. Optimizing quantum networks: the method of quantum combs

A convenient way to optimize quantum networks is the method of *quantum combs* [25, 26] (see also the work by Gutoski and Watrous [27]). Here we briefly summarize some known basic facts about this method, referring the reader to the original papers for the proofs and for further details.

In the following we will use the following notation: $\text{Lin}(\mathcal{H})$ will denote the set of linear operators on a (finite-dimensional) Hilbert space \mathcal{H} , $\text{Lin}_+(\mathcal{H})$ will denote the set of positive operators on \mathcal{H} , while $\text{St}(\mathcal{H})$ will denote the set of density matrices on \mathcal{H} , that is the set of positive operators $\rho \in \text{Lin}_+(\mathcal{H})$ such that $\text{Tr}[\rho] = 1$.

3.1. Quantum combs.

A network of quantum channels with internal memories can be associated with a non-negative operator satisfying suitable linear constraints. Precisely, a network of the form



is associated to a positive operator $R \in \text{Lin}_+ \left[\bigotimes_{s \in \mathcal{S}} \left(\mathcal{H}_{out}^{(s)} \otimes \mathcal{H}_{in}^{(s)} \right) \right]$. The fact that the network consists of quantum channels (trace-preserving maps) imposes the following constraint: there must exist a set of positive operators $R^{(n)} \in$

$$\text{Lin}_+ \left[\bigotimes_{i=1}^n \left(\mathcal{H}_{out}^{(s_i)} \otimes \mathcal{H}_{in}^{(s_i)} \right) \right], \quad n = 1, \dots, N-1 \text{ such that}$$

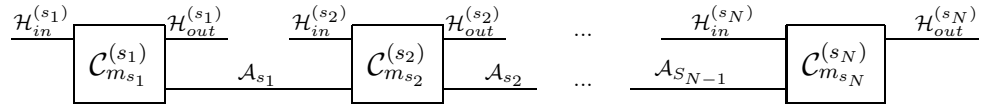
$$\begin{cases} \text{Tr}_{out, s_N} [R] &= I_{in, s_N} \otimes R^{(N-1)} \\ \text{Tr}_{out, s_{N-1}} [R^{(N-1)}] &= I_{in, s_{N-1}} \otimes R^{(N-2)} \\ &\vdots \\ \text{Tr}_{out, s_1} [R^{(1)}] &= I_{in, s_1}, \end{cases} \quad (3)$$

where $\text{Tr}_{out, s}$ and $I_{in, s}$ denote the partial trace over $\mathcal{H}_{out}^{(s)}$ and the identity operator on $\mathcal{H}_{in}^{(s)}$ [27, 25, 26].

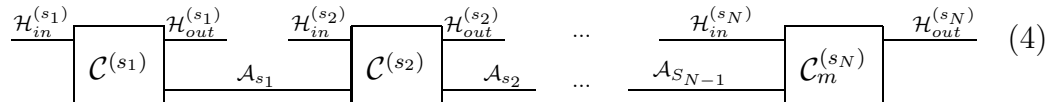
Most importantly, the converse also holds [27, 25, 26]: if a positive operator R satisfies the constraints of Eq. (3) for some set of positive operators $R^{(n)}, n = 1, \dots, N-1$, then there exists a network of the form of Eq. (2) such that the operator associated to that network is R . This is important because it implies that optimizing over quantum networks is completely equivalent to optimizing over positive operators R satisfying Eq. (3). In fact, given an operator R satisfying there is a constructive algorithm to build up the channels $\mathcal{C}^{(s)}$ at all time steps $s \in S$ [29]. In the following, a positive operator $R \in \text{Lin}_+ \left[\bigotimes_{s \in S} \left(\mathcal{H}_{out}^{(s)} \otimes \mathcal{H}_{in}^{(s)} \right) \right]$ satisfying Eq. (3) for some operators $R^{(n)}, n = 1, \dots, N-1$ will be called *quantum comb*. We will denote the set of quantum combs as $\text{Comb} \left[\bigotimes_{s \in S} \left(\mathcal{H}_{out}^{(s)} \otimes \mathcal{H}_{in}^{(s)} \right) \right]$.

3.2. Quantum testers.

More generally, a quantum network can contain measurements: at each time step s one can have a measurement with outcome m_s in some set \mathbf{M}_s . Conditionally to the outcome m_s , the input system will undergo a random transformation, represented by a completely positive trace non-increasing map $\mathcal{C}_{m_s}^{(s)}$, with the condition that the sum over all outcomes $\mathcal{C}^{(s)} := \sum_{m_s \in \mathbf{M}_s} \mathcal{C}_{m_s}^{(s)}$ is trace-preserving. A network containing measurements, such as the network



can be associated with a collection of positive operators $\mathbf{T} := \{T_m \mid m \in \mathbf{M} := \mathbf{M}_1 \times \dots \times \mathbf{M}_N\}$ with the property that the sum over all outcomes $T := \sum_{m \in \mathbf{M}} T_m$ satisfies Eq. (3). We call such a collection of operators a *quantum tester*. It is possible to prove that, if a collection positive operators $\mathbf{T} = \{T_m \mid m \in \mathbf{M}\}$ is a quantum tester, then there exists a quantum network of the form



such that \mathbf{T} is the tester associated to that network [27, 25, 26]. Note that here the measurement takes place only in the last step, while the boxes $\mathcal{C}^{(s_n)}, n = 1, \dots, N-1$ represent quantum channels.

A particular type of testers are those where the first and last quantum systems are trivial [$\mathcal{H}_{in}^{(s_1)} \simeq \mathcal{H}_{out}^{(s_N)} \simeq \mathbb{C}$ in Eq. (4)]. These testers represent quantum networks that start with a state preparation and end with a POVM measurement. These are exactly the networks that are interesting for the estimation of quantum processes, as depicted in Eq. (1): note that to test a process consisting of N time steps we need tester consisting of $N + 1$ time steps. Labelling the Hilbert spaces as in the following diagram

$$\begin{array}{c} \text{---} \mathcal{H}_{in}^{(s_1)} \text{---} \mathcal{H}_{out}^{(s_1)} \text{---} \mathcal{H}_{in}^{(s_2)} \text{---} \dots \text{---} \mathcal{H}_{out}^{(s_N)} \text{---} \\ \text{---} \mathcal{C}^{(s_1)} \text{---} \mathcal{A}_{s_1} \text{---} \mathcal{C}^{(s_2)} \text{---} \mathcal{A}_{s_2} \text{---} \dots \text{---} \mathcal{A}_{s_N} \text{---} \mathcal{C}_m^{(s_{N+1})} \end{array} \quad (5)$$

the normalization of the tester \mathbf{T} becomes

$$\left\{ \begin{array}{l} \sum_{m \in \mathbf{M}} T_m = I_{out, s_N} \otimes \Xi^{(N)} \\ \text{Tr}_{in, s_N} [\Xi^{(N)}] = I_{out, s_{N-1}} \otimes \Xi^{(N-1)} \\ \vdots \\ \text{Tr}_{in, s_1} [\Xi^{(1)}] = 1, \end{array} \right. \quad (6)$$

for some set of positive operators $\Xi^{(n)} \in \text{Lin} \left\{ \mathcal{H}_{in}^{(s_n)} \otimes \left[\bigotimes_{i=1}^{n-1} \left(\mathcal{H}_{out}^{(s_i)} \otimes \mathcal{H}_{in}^{(s_i)} \right) \right] \right\}$, $n = 1, \dots, N$.

3.3. Generalized Born rule.

If we test a process represented by the quantum comb $R \in \text{Comb} \left[\bigotimes_{s \in \mathbf{S}} \left(\mathcal{H}_{out}^{(s)} \otimes \mathcal{H}_{in}^{(s)} \right) \right]$ with a network represented by the tester $\mathbf{T} := \{T_m \mid m \in \mathbf{M}\}$, then we obtain a probability distribution $p(m|R^{(N)})$ over all possible outcomes, given by the *generalized Born rule* [25, 26]

$$p(m|R) = \text{Tr} [T_m R]. \quad (7)$$

Here the quantum comb R plays the role of the density matrix in the ordinary Born rule, and the tester $\{T_m \mid m \in \mathbf{M}\}$ plays the role of the POVM measurement. In fact, the ordinary Born rule is a very special case of Eq. (7), corresponding to the case of *state preparation processes*, which consist of a single time step ($N = 1$) with no input system ($\mathcal{H}_{in}^{(s_1)} \simeq \mathbb{C}$).

4. The optimization problem of Quantum Metrology

In process estimation one has a parametric family of processes with a given input-output structure and with a fixed number of time steps N labelled by an index $s \in \mathbf{S} \subset \mathbb{N}$. Each process is described by a quantum comb $R_x \in \text{Comb} \left[\bigotimes_{s \in \mathbf{S}} \left(\mathcal{H}_{out}^{(s)} \otimes \mathcal{H}_{in}^{(s)} \right) \right]$, where $x \in \mathbf{X}$ is the parameter to be estimated. Let us denote by $\pi(x)$ the probability that the unknown parameter has the value x . If x has a continuum of values, $p(x)$ will represent the probability density of x with respect to some measure \mathfrak{x} . For simplicity in the following we will present the results in the discrete case, but it is important to bear in mind that these results hold also in the continuous case, just replacing sums with integrals and replacing the quantifier “ $\forall \hat{x} \in \mathbf{X}$ ” with “ $\forall \hat{x} \in \mathbf{X}$ except at most for a set of zero measure”.

4.1. Primal maximization problem

For an estimation strategy described by the quantum tester $\mathbf{T} := \{T_{\hat{x}} \mid \hat{x} \in \mathbf{X}\}$, the probability distribution $p(\hat{x}|x)$ is given by Eq. (7). In order to evaluate the performance of a given strategy, we introduce a payoff function $g(\hat{x}, x)$, which quantifies the gain (or the loss) obtained by estimating \hat{x} when the actual value is x . In general, as long as the payoff is lower bounded (that is, as long as there is a limit to the losses) one can choose without loss of generality $g(\hat{x}, x) \geq 0$, $\forall \hat{x}, x \in \mathbf{X}$. The expected payoff, averaged over the possible true values, is then given by

$$\begin{aligned} \gamma[\mathbf{T}] &:= \sum_{x \in \mathbf{X}} \pi(x) \sum_{\hat{x} \in \mathbf{X}} g(\hat{x}, x) p(\hat{x}|x) \\ &= \sum_{\hat{x} \in \mathbf{X}} \text{Tr}[T_{\hat{x}} G_{\hat{x}}] \quad G_{\hat{x}} = \sum_{x \in \mathbf{X}} \pi(x) g(\hat{x}, x) R_x. \end{aligned} \quad (8)$$

An example of payoff function is $g(\hat{x}, x) = \delta_{\hat{x}, x}$, which gives a unit gain if and only if the estimated value \hat{x} coincides with the true value x . In this case the average gain coincides with the average probability of guessing the correct value

$$\gamma[\mathbf{T}] \equiv p_{\text{succ}} := \sum_{x \in \mathbf{X}} \pi(x) p(x|x).$$

A tester \mathbf{T} is optimal if it achieves the maximum payoff, defined as

$$\begin{aligned} \gamma_{\max} &:= \max_{\mathbf{T}, \Xi^{(1)}, \dots, \Xi^{(N)}} \gamma[\mathbf{T}] \\ &\quad T_{\hat{x}} \geq 0, \forall \hat{x} \in \mathbf{X} \\ &\quad \sum_{\hat{x} \in \mathbf{X}} T_{\hat{x}} = I_{\text{out}, s_N} \otimes \Xi^{(N)} \\ &\quad \text{Tr}_{\text{in}, s_N} [\Xi^{(N)}] = I_{\text{out}, s_{N-1}} \otimes \Xi^{(N-1)} \\ &\quad \vdots \\ &\quad \text{Tr}_{\text{in}, s_1} [\Xi^{(1)}] = 1. \end{aligned}$$

4.2. Dual minimization problem

Maximizing the payoff in Eq. (9) is a semidefinite program. Using duality theory we now give a useful expression for the maximum payoff:

Theorem 1 *The maximum payoff is given by*

$$\gamma_{\max} = \min \left\{ \lambda \geq 0 \mid \exists R \in \text{Comb} \left[\bigotimes_{s \in \mathbf{S}} \left(\mathcal{H}_{\text{out}}^{(s)} \otimes \mathcal{H}_{\text{in}}^{(s)} \right) \right] : \lambda R \geq G_{\hat{x}}, \quad \forall \hat{x} \in \mathbf{X} \right\}, \quad (9)$$

where $G_{\hat{x}}$ is defined as in Eq. (8).

The proof of the theorem, given in the Appendix, follows the same lines used by Gutoski [35] to prove strong duality for the minimum error discrimination of two quantum processes, which is the special instance of our problem corresponding to $\mathbf{X} := \{0, 1\}$ and $g(\hat{x}, x) = \delta_{\hat{x}, x}$. Here we illustrate the result of theorem 1 in a few special examples.

4.3. Examples

4.3.1. State estimation. State estimation can be viewed as a special case where the unknown process \mathcal{P}_x to be estimated consists only in the preparation of a quantum state $\rho_x \in \text{Lin}_+(\mathcal{H})$ (that is, when there is only one time step $N = 1$, the output Hilbert space is $\mathcal{H}_{out}^{(s_1)} = \mathcal{H}$, and the input Hilbert space is trivial $\mathcal{H}_{in}^{(s_1)} \simeq \mathbb{C}$). In this case, the expression (9) becomes

$$\gamma_{\max} = \min \{ \lambda \geq 0 \mid \exists \rho \in \text{St}(\mathcal{H}) : \lambda \rho \geq G_{\hat{x}}, \quad \forall \hat{x} \in \mathbf{X} \}, \quad (10)$$

with $G_{\hat{x}} = \sum_{x \in \mathbf{X}} \pi(x) g(\hat{x}, x) \rho_x$.

4.3.2. Minimum error state discrimination. If $g(\hat{x}, x) = \delta_{\hat{x}, x}$, the maximum payoff γ_{\max} coincides with the maximum probability of guessing the correct value p_{succ}^{\max} , so that maximizing the payoff is equivalent to minimizing the error probability. In this special case we retrieve from Eq. (10) the classic expression by Yuen, Kennedy, and Lax [28] (see also [30, 31])

$$p_{\text{succ}}^{\max} = \min \{ \text{Tr}[\Lambda] \mid \Lambda \in \text{Lin}(\mathcal{H}), \quad \Lambda \geq \pi_{\hat{x}} \rho_{\hat{x}}, \forall \hat{x} \in \mathbf{X} \} \quad (11)$$

[the above expression follows from Eq. (10) with the definition $\Lambda := \lambda \rho$].

4.3.3. State estimation/discrimination in the group covariant case. The dual expression for the maximum payoff has an interesting interpretation in the presence of symmetry. Let us first consider a simple case of state discrimination, where \mathbf{X} is a finite group, the prior probability π is uniform, that is, $\pi(x) = 1/|\mathbf{X}|$, and the unknown state ρ_x is given by $\rho_x = U_x \rho_0 U_x^\dagger$, where $\rho_0 \in \text{St}(\mathcal{H})$ is a fixed state and $U : \mathbf{X} \rightarrow \text{Lin}(\mathcal{H}), x \mapsto U_x$ is a unitary representation of the group \mathbf{X} . In this case, it is easy to show that the minimization over $\Lambda = \lambda \rho$ in Eq. (11) can be restricted without loss of generality to invariant states, satisfying $U_x \rho U_x^\dagger = \rho, \forall x \in \mathbf{X}$. Hence, we have

$$\begin{aligned} p_{\text{succ}}^{\max} &= \min \left\{ \lambda \mid \exists \rho \in \text{St}(\mathcal{H}) : \rho \text{ is invariant}, \rho \geq \frac{\rho_0}{\lambda |\mathbf{X}|} \right\} \\ &= \frac{1}{|\mathbf{X}| q_{\max}} \\ q_{\max} &:= \max \{ q \mid \exists \rho \in \text{St}(\mathcal{H}) : \rho \text{ is invariant}, q \rho_0 \leq \rho \} \end{aligned} \quad (12)$$

By definition, q_{\max} is the maximum probability that ρ_0 can have in an ensemble decomposition of an invariant state ρ , optimized over all possible invariant states. The probability q_{\max} ranges between $1/|\mathbf{X}|$ and 1. Intuitively, it can be interpreted as a measure of how symmetric is the state ρ_0 : for $q_{\max} = 1$ the state ρ_0 is invariant, while for $q_{\max} = 1/|\mathbf{X}|$ the state ρ_0 generates a family of orthogonal states $\rho_x = U_x \rho_0 U_x^\dagger$.

The result can be easily extended to the case of arbitrary payoff functions that are *left-invariant* under the action of the group, that is, functions g satisfying the condition $g(y\hat{x}, yx) = g(\hat{x}, x), \forall \hat{x}, x, y \in \mathbf{X}$. Moreover, the expression of Eq. (12) can be generalized to a form that holds also for continuous groups:

Corollary 1 *Let X be a compact group, $g : \mathsf{X} \times \mathsf{X} \rightarrow \mathbb{R}$ be a left-invariant payoff function, and ρ_x be the quantum state $\rho_x := U_x \rho_0 U_x^\dagger$, where $U : x \mapsto U_x$ is a unitary representation of the group X . If the prior probability is given by the Haar measure x , then the maximum average payoff over all quantum measurements is given by*

$$\begin{aligned} \gamma_{\max} &= \frac{\gamma_0}{q_{\max}} & \gamma_0 &:= \int_{\mathsf{X}} x \, g(e, x) \\ q_{\max} &:= \max \{q \mid \exists \rho \in \text{St}(\mathcal{H}) : \rho \text{ is invariant, } q\sigma_0 \leq \rho\} \\ \sigma_0 &:= \frac{1}{\gamma_0} \int_{\mathsf{X}} x \, g(e, x) \, U_x \rho_0 U_x^\dagger, \end{aligned}$$

where $e \in \mathsf{X}$ denotes the identity element in the group X .

Proof. Using the invariance of the Haar measure and of the payoff function it is easy to check that $G_{\hat{x}} = U_{\hat{x}}(\gamma_0 \sigma_0) U_{\hat{x}}^\dagger$. Using this fact, we can restrict the minimization in Eq. (10) to invariant states ρ satisfying the condition $\lambda \rho \geq \gamma_0 \sigma_0$. Finally, defining $q := \gamma_0 / \lambda$ we can transform the minimization over λ into a maximization over q , thus proving the thesis. ■

4.3.4. Binary discrimination of multi-time quantum processes The discrimination of two multi-time processes \mathcal{P}_0 and \mathcal{P}_1 corresponds to the special case where $\mathsf{X} = \{0, 1\}$. In this case, the maximum probability of successful discrimination defines an operational norm in the real vector space generated by quantum processes [34, 35]. For prior probabilities π_0 and π_1 , the probability of success and the norm are linked by the relation [34]

$$p_{\text{succ}} = \frac{1}{2} (1 + \|\pi_0 \mathcal{P}_0 - \pi_1 \mathcal{P}_1\|_{\text{op}}),$$

which generalizes the well-known expression by Helstrom [16] for the optimal discrimination between two quantum states. In the binary case the dual expression for the maximum success probability given by theorem 1 coincides with the dual expression presented by Gutoski in Ref. [35].

4.3.5. Process estimation/discrimination in the group covariant case. Consider the case of a general process \mathcal{P}_x consisting of N time steps. Suppose that \mathcal{P}_x has the form $\mathcal{P}_x = \left(\bigotimes_{s \in \mathsf{S}} \mathcal{V}_x^{(s)} \right) \mathcal{P}_0 \left(\bigotimes_{s \in \mathsf{S}} \mathcal{U}_x^{(s)\dagger} \right)$, where \mathcal{P}_0 is a fixed process and $\mathcal{U}_x^{(s)\dagger}(\rho) := U_x^{(s)\dagger} \rho U_x^{(s)} \left[\mathcal{V}_x^{(s)}(\rho) := V_x^{(s)} \rho V_x^{(s)\dagger} \right]$ is a unitary quantum channel representing the action of the group on the input (output) system at the s -th time step.

Denoting by R_x and R_0 the quantum combs corresponding to the processes \mathcal{P}_x and \mathcal{P}_0 , it is possible to show that $R_x = \left(\bigotimes_{s \in \mathsf{S}} \mathcal{V}_x^{(s)} \otimes \mathcal{U}_x^{(s)*} \right) (R_0)$ where $\mathcal{U}_x^{(s)*}$ denotes the complex conjugate $\mathcal{U}_x^{(s)*}$ with respect the computational basis [32].

The result of Corollary 1 can then be generalized immediately to the case of general processes:

Corollary 2 Let X be a compact group, $g : \mathsf{X} \times \mathsf{X} \rightarrow \mathbb{R}$ be a left-invariant payoff function, and let ρ_x be the quantum state $\rho_x := U_x \rho_0 U_x^\dagger$, where $U : x \mapsto U_x$ is a unitary representation of the group X . If the prior probability is given by the Haar measure \mathfrak{x} , then the maximum average payoff over all quantum measurements is given by

$$\begin{aligned} \gamma_{\max} &= \frac{\gamma_0}{q_{\max}} \\ \gamma_0 &:= \int_{\mathsf{X}} \mathfrak{x} \, g(e, x) \\ q_{\max} &:= \max \left\{ q \mid \exists R \in \text{Comb} \left(\bigotimes_{s \in \mathsf{S}} \mathcal{H}_{\text{out}}^{(s)} \otimes \mathcal{H}_{\text{in}}^{(s)} \right) : R \text{ is invariant, } q S_0 \leq R \right\} \\ S_0 &:= \frac{1}{\gamma_0} \int_{\mathsf{X}} \mathfrak{x} \, g(e, x) \left(\bigotimes_{s \in \mathsf{S}} \mathcal{V}_x^{(s)} \otimes \mathcal{U}_x^{(s)*} \right) (R_0), \end{aligned}$$

where $e \in \mathsf{X}$ denotes the identity element in the group X .

Proof. Same proof as for corollary 1. ■

5. Product rule for the estimation of independent processes

Imagine that we have K processes, where each process \mathcal{P}_{k,x_k} corresponds to a quantum network as in figure (2) and is labelled by an unknown parameter x_k in some set X_k , $k = 1, \dots, K$. For every fixed k , all the processes $\{\mathcal{P}_{k,x_k} \mid x_k \in \mathsf{X}_k\}$ consist of the same number N_k of time steps, which we label by an index s_k in some set $\mathsf{S}_k \subset \mathbb{N}$. At time s_k , each process \mathcal{P}_{k,x_k} will transform an input system with Hilbert space $\mathcal{H}_{k,\text{in}}^{(s_k)}$, into an output system with Hilbert space $\mathcal{H}_{k,\text{out}}^{(s_k)}$.

Let us denote by \mathbf{x} the vectors of parameters $\mathbf{x} := (x_1, \dots, x_K) \in \mathsf{X} := \mathsf{X}_1 \times \dots \times \mathsf{X}_K$. We say that the K processes $\{\mathcal{P}_{k,x_k} \mid k = 1, \dots, K\}$ are *independent* when

- two processes $\mathcal{P}_{k,x_k} = \mathcal{P}_{l,x_l}$ with $k \neq l$ correspond to two disconnected quantum networks for every $x_k \in \mathsf{X}_k$ and for every $x_l \in \mathsf{X}_l$
- the prior distribution of the parameters factorizes as

$$\pi(\mathbf{x}) = \pi_1(x_1) \pi_2(x_2) \cdots \pi_K(x_K), \quad (13)$$

where π_k is the prior distribution for the parameter x_k .

For example, the different parameters could be K independent and uniformly distributed phase shifts.

If $\{\mathcal{P}_{k,x_k} \mid k = 1, \dots, K\}$ are K independent processes, we denote by $\mathcal{P}_{\mathbf{x}} := \mathcal{P}_{1,x_1} \otimes \mathcal{P}_{2,x_2} \otimes \dots \otimes \mathcal{P}_{K,x_K}$ the corresponding joint process.

Suppose that we want to estimate parameter \mathbf{x} labelling the joint process $\mathcal{P}_{\mathbf{x}}$ and that our figure of merit is given by the payoff function $g(\hat{\mathbf{x}}, \mathbf{x})$. If we are interested in

each parameter independently, then the payoff function for the estimation of the vector \mathbf{x} is the product of the payoff functions for the estimation of its components:

$$g(\hat{\mathbf{x}}, \mathbf{x}) = \prod_{k=1}^K g_k(\hat{x}_k, x_k) \quad g_k \geq 0, \forall k = 1, \dots, K, \quad (14)$$

where the notation $g_k \geq 0$ means $g(\hat{x}_k, x_k) \geq 0, \forall \hat{x}_k, x_k \in \mathbf{X}_k$. For example, the payoff function could give a reward only when all the parameters are guessed correctly, so that $g(\hat{\mathbf{x}}, \mathbf{x}) = \delta_{\hat{\mathbf{x}}, \mathbf{x}} = \prod_{k=1}^K \delta_{\hat{x}_k, x_k}$.

Note that, in order to have a meaningful figure of merit for the estimation of the vector \mathbf{x} , it is important to have $g_n \geq 0$ for every n : otherwise, the product of two negative gains (i.e. of two losses) for two different parameters would count as a positive gain for the joint estimation of the vector \mathbf{x} .

Based on the hypotheses of independence of the processes and on the product form of the payoff function we can prove the following theorem:

Theorem 2 (Product rule for the estimation of K independent processes)

Let $\mathcal{P}_{k, x_k}, k = 1, \dots, K$ be K independent processes, each process labelled by an unknown parameter $x_k \in \mathbf{X}_k$ with prior probability $\pi_k(x_k)$. Then for a payoff function $g(\hat{\mathbf{x}}, \mathbf{x})$ of the product form of Eq. (14) the maximum payoff for the estimation of \mathbf{x} is given by the product of the maximum payoffs for the estimation of its components:

$$\gamma_{\max} = \prod_{k=1}^K \gamma_{\max, k}, \quad (15)$$

where $\gamma_{\max, k}$ is the maximum payoff achievable in the estimation of x_k .

In other words, the optimal estimation of the vector \mathbf{x} can be achieved by estimating each component x_k independently.

Proof. Clearly, we have $\gamma_{\max} \geq \prod_{k=1}^K \gamma_{\max, k}$, because restricting to product strategies can only reduce the maximum payoff. To prove the converse we use the dual minimization problem of Theorem 1, in which restricting to product combs can only increase the minimum.

Let R_{k, x_k} be the quantum comb representing the process \mathcal{P}_{k, x_k} and let $R_{\mathbf{x}} = \bigotimes_{k=1}^K R_{k, x_k}$ be the quantum comb representing the process $\mathcal{P}_{\mathbf{x}} = \bigotimes_{k=1}^K \mathcal{P}_{k, x_k}$. Let us introduce the notation

$$\begin{aligned} \mathbf{C}_k &:= \text{Comb} \left[\left(\bigotimes_{s_k \in \mathbf{S}_k} \mathcal{H}_{\text{out}}^{(s_k)} \otimes \mathcal{H}_{\text{in}}^{(s_k)} \right) \right] \\ \mathbf{C} &:= \text{Comb} \left[\left(\bigotimes_{k=1}^K \bigotimes_{s_k \in \mathbf{S}_k} \mathcal{H}_{\text{out}}^{(s_k)} \otimes \mathcal{H}_{\text{in}}^{(s_k)} \right) \right] \\ \mathbf{C}_{\text{prod}} &:= \left\{ R = \bigotimes_{k=1}^K R_k \mid R_k \in \mathbf{C}_k \ \forall k = 1, \dots, K \right\} \subset \mathbf{C}. \end{aligned}$$

With this notation we have that R_{k, x_k} and $R_{\mathbf{x}}$ belong to \mathbf{C}_k and \mathbf{C} , respectively.

Define the positive operators

$$G_{k,\hat{x}_k} := \sum_{x_k \in \mathbf{X}_k} \pi_k(x_k) g_k(\hat{x}_k, x_k) R_{k,x_k}$$

$$G_{\hat{\mathbf{x}}} := \sum_{\mathbf{x} \in \mathbf{X}} \pi(\mathbf{x}) g(\hat{\mathbf{x}}, \mathbf{x}) R_{\mathbf{x}} \equiv \bigotimes_{k=1}^K G_{k,\hat{x}_k}.$$

Then, by theorem 1 we have

$$\begin{aligned} \gamma_{\max} &= \min \{ \lambda \geq 0 \mid \exists R \in \mathbf{C} : \lambda R \geq G_{\mathbf{x}}, \quad \forall \mathbf{x} \in \mathbf{X} \} \\ &\leq \min \{ \lambda \geq 0 \mid \exists R \in \mathbf{C}_{\text{prod}} : \lambda R \geq G_{\mathbf{x}}, \quad \forall \mathbf{x} \in \mathbf{X} \} \\ &\leq \prod_{k=1}^K \min \{ \lambda_k \geq 0 \mid \exists R_k \in \mathbf{C}_k : \lambda_k R_k \geq G_{k,x_k}, \quad \forall x_k \in \mathbf{X}_k \} \\ &= \prod_{k=1}^K \gamma_{\max,k}. \end{aligned}$$

Here, the second inequality comes from the fact that if $\lambda_k R_k \geq G_{k,x_k}$ for all k , then $\lambda R \geq G_{\mathbf{x}}$ for $\lambda = \prod_k \lambda_k$ and $R = \bigotimes_k R_k$. ■

5.0.6. Relation with the product rules by Mittal and Szegedy. The technique used to prove that the optimal payoff is of the product form is directly inspired by a result by Mittal and Szegedy on product rules for semidefinite programming [23]. However, our result is not a direct application of the theorem in Ref. [23], which concerns *product programs*, where the linear constraint for the product program is the tensor product of the linear constraints for the individual programs. The theorem is not directly applicable in our case because in the joint estimation of K processes the linear constraint of Eq. (9) are not the tensor product of the linear constraints for the estimation each process separately. However, the crucial point here is that the tensor product of K operators satisfying the constraints individually is an operator that satisfies the joint constraint and that this property is true both in the primal maximization problem and in the dual minimization program.

5.0.7. Example 5: minimum error discrimination of K sets of processes Theorem 2 can be applied to the case of minimum error discrimination of processes. Suppose that for every $k = 1, \dots, K$ we have a set of processes $\{\mathcal{P}_{k,x_k} \mid x_k \in \mathbf{X}_k\}$, each process \mathcal{P}_{k,x_k} having prior probability π_{k,x_k} ($\sum_{x_k \in \mathbf{X}_k} \pi_{k,x_k} = 1$). Denoting by $p_{\text{succ},k}^{\max}$ the maximum probability of success in correctly identifying the k -th process, and by p_{succ}^{\max} the probability of success in correctly identifying all processes, we then have $p^{\max} = p_{\text{succ},1}^{\max} \cdots p_{\text{succ},K}^{\max}$. The best joint strategy for discrimination is just the product of the best individual strategies.

5.1. Counterexamples

Our theorem 2 proved the optimality of product strategies in the hypotheses that the processes are independent and that the payoff function is of a product form. Here we

show that if these hypotheses are dropped, the result may not hold.

5.1.1. Minimum error discrimination of two pure states with multiple copies. Consider the minimum error discrimination of two pure states $\{\rho_0, \rho\}$ with prior probabilities $\{p_0, p_1\}$, in the case where K identical copies of the unknown state are available. We can view this problem as an instance of minimum error discrimination of K perfectly correlated preparation processes, each of which prepares one of the states $\{\rho_0, \rho_1\}$. Clearly, denoting by $p_{succ}^{\max}(K)$ the probability of success with K copies, we have that $p_{succ}^{\max}(K)$ converges to 1 exponentially fast in the limit $K \rightarrow \infty$ [33]. On the other hand, the product of the probabilities of success, given by $[p_{succ}^{\max}(K=1)]^K$ tends to zero (exponentially fast) unless the two states are perfectly distinguishable.

5.1.2. Estimation of two independent phase shifts with a correlated payoff function. Consider the estimation of two independent phase shifts on two qubit systems, with Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 , respectively ($\mathcal{H}_1 \simeq \mathcal{H}_2 \simeq \mathbb{C}^2$). Denoting by $|0\rangle$ and $|1\rangle$ the two orthonormal vectors in the standard basis in \mathbb{C}^2 , the phase shifts on a qubit system are given by $U_x = |0\rangle\langle 0| + e^{ix}|1\rangle\langle 1|$, $x \in [0, 2\pi)$. We assume that the phase shifts on the two qubits are uniformly distributed according to the Haar measure $x/2\pi$. The problem is then to find the best estimate of the unknown parameter $\mathbf{x} := (x_1, x_2)$ characterizing the black boxes U_{x_1} and U_{x_2} . As a figure of merit, we consider the maximization of the payoff function

$$g_p(\hat{\mathbf{x}}, \mathbf{x}) = p \cos(\hat{x}_1 + \hat{x}_2 - x_1 - x_2) + (1 - p) \cos(\hat{x}_1 - \hat{x}_2 - x_1 + x_2),$$

for some $p \in [0, 1]$. Note that g_p is a convex combination of the figure of merit $\cos(\hat{x}_1 + \hat{x}_2 - x_1 - x_2)$, which quantifies how good is our estimate of the sum $s := x_1 + x_2$, and of the figure of merit $\cos(\hat{x}_1 - \hat{x}_2 - x_1 + x_2)$, which quantifies how good is our estimate to compute the difference $d := x_1 - x_2$. In other words, we can interpret f as expressing the fact that, with probability p , we will be asked to estimate the sum, while with probability $(1 - p)$ we will be asked to estimate the difference.

Due to the symmetry of the problem, it is enough to consider quantum networks where the two unknown phase shifts are applied in parallel on a suitable entangled state $|E\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$, as proven in Ref. [34]. No additional reference system is needed, because the black boxes form a unitary representation of an abelian group [36]. Hence, the problem is reduced to the optimal estimation of \mathbf{x} from the output state $|E_{\mathbf{x}}\rangle := (U_{x_1} \otimes U_{x_2})|E\rangle$.

From the theory of optimal estimation of group parameters [36] we know that the optimal measurement is given by the covariant POVM

$$P_{\hat{\mathbf{x}}} = (U_{x_1} \otimes U_{x_2})|\eta\rangle\langle\eta|(U_{x_1} \otimes U_{x_2})^\dagger \quad |\eta\rangle := |0\rangle|0\rangle + |0\rangle|1\rangle + |1\rangle|0\rangle + |1\rangle|1\rangle.$$

Incidentally, we note that the POVM is of the product form $P_{\hat{\mathbf{x}}} = P_{1, \hat{x}_1} \otimes P_{2, \hat{x}_2}$. By direct calculation, we then find that the average value of g_p is $\gamma_p = \langle E|G_p|E\rangle$ with

$$G_p = \frac{p}{2}(|0\rangle|0\rangle\langle 1|\langle 1| + |1\rangle|1\rangle\langle 0|\langle 0|) + \frac{1-p}{2}(|0\rangle|1\rangle\langle 1|\langle 0| + |1\rangle|0\rangle\langle 0|\langle 1|).$$

Clearly, the maximum eigenvalue of G_p is $\lambda_{\max} = \max\{p/2, (1-p)/2\}$, corresponding to the nondegenerate eigenvector $|E\rangle = 2^{-\frac{1}{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$ for $p > 1/2$ and $|E\rangle = 2^{-\frac{1}{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle)$ for $p < 1/2$. For $p = 1/2$ one has degeneration, and the optimal input state can be chosen of the product form $|E\rangle = |+\rangle|+\rangle$ with $|+\rangle = 2^{-\frac{1}{2}}(|0\rangle + |1\rangle)$.

The qualitative explanation of the behaviour is the following: For $p = 1/2$ the figure of merit is factorized ($g_{\frac{1}{2}} = \cos(\hat{\varphi} - \varphi) \cos(\hat{\psi} - \psi)$) and the estimation strategy is factorized too. For every value $p \neq \frac{1}{2}$, the degeneration is removed and suddenly the optimal input state becomes maximally entangled. Note, however, that there is no discontinuity in the average payoff.

5.1.3. Estimating the sum of K independent phase shifts. Suppose that we have K identical systems, with Hilbert spaces $\mathcal{H}_k \simeq \mathbb{C}^N$ for all $k = 1, \dots, K$, and suppose that each system undergoes an independent phase shift $U_{x_k}^{(k)} := e^{ix_k H^{(k)}}$, where $H^{(k)} := \sum_{n=1}^N n |n\rangle\langle n|$ for every k , $\{|n\rangle\}$ being the computational basis.

If we want to estimate the sum $s := \sum_k x_k$ a natural figure of merit is the cost function $c(\hat{s}, s) = 2[1 - \cos(\hat{s} - s)]$. This cost function is well known in the phase estimation literature as a smooth and periodic version of the variance [16, 17, 4, 12]. For small s , we have indeed $\hat{c}(\hat{s}, s) \approx (\hat{s} - s)^2$. Clearly, minimizing c is equivalent to maximizing the payoff function $g(\hat{s}, s) = 1 + \cos(\hat{s} - s)$.

Let us find the optimal estimation strategy. First, using the fact that the unknown black boxes form a unitary representation of an abelian group, we know that the optimal strategy consists in applying the black boxes in parallel on an entangled input state $|E\rangle \in \mathcal{H}^{\otimes K}$. Moreover, note that for every fixed i and j , if we apply the transformation $x_i \mapsto x_i + \xi$, $\hat{x}_i \mapsto \hat{x}_i + \xi$, $x_j \mapsto x_j - \xi$, $\hat{x}_j \mapsto \hat{x}_j - \xi$, $\xi \in [0, 2\pi)$, then the value of the figure of merit does not change. Using this symmetry it is easy to show that the input state $|E\rangle$ must be an eigenstate of the difference operator $\Delta_{ij} = H^{(i)} - H^{(j)}$ for every possible pair i, j . It is then straightforward that the optimal choice is $|E\rangle = \sum_{n=1}^N e_n |n\rangle^{\otimes K}$, where $\{e_n\}$ are suitable coefficients. The problem then becomes to estimate the sum s from the state $|E_{\mathbf{x}}\rangle := \left(\prod_k U_{x_k}^{(k)}\right) |E\rangle = \sum_{n=1}^N e^{isn} e_n |n\rangle^{\otimes K}$. From the theory of optimal phase estimation we know that the minimum cost is $c_{\min} = 4 \sin^2 \left[\frac{\pi}{2N} \right]$, which converges to $\frac{\pi^2}{N^2}$ in the limit $N \rightarrow \infty$ (see Ref. [4]). The corresponding optimal state is the entangled state [4]

$$|E_{\text{opt}}\rangle = \left(\frac{N}{2}\right)^{-\frac{1}{2}} \sum_{n=1}^N \sin \left[\frac{\pi(n-1)}{(N-1)} \right] |n\rangle^{\otimes K}.$$

and the optimal POVM is $P_s = |\eta_s\rangle\langle\eta_s|$, $|\eta_s\rangle := \sum_{n=1}^N e^{isn} |n\rangle^{\otimes K}$. It is easy to see that the use of entanglement implies an advantage over factorized strategies, where each system is prepared independently in a state $|e_k\rangle$ and is measured independently with the optimal POVM. Indeed, if we choose the optimal states $|e_k\rangle = |e\rangle := \left(\frac{N}{2}\right)^{-\frac{1}{2}} \sum_{n=1}^N \sin \left[\frac{\pi(n-1)}{N-1} \right]$ and the optimal product POVM $P_{\hat{\mathbf{x}}} := \prod_k U_{x_k}^{(k)} (2|+\rangle\langle+|) U_{x_k}^{(k)\dagger}$ then we obtain the cost

$$\langle c(\hat{s}, s) \rangle = 2(1 - \langle \cos(\hat{s} - s) \rangle)$$

$$\begin{aligned}
&= 2 \left(1 - \prod_{k=1}^K \langle \cos(\hat{x}_k - x_k) \rangle \right) \\
&= 2 \left\{ 1 - \left[1 - 2 \sin^2 \left(\frac{\pi}{2M} \right) \right]^K \right\},
\end{aligned}$$

where $\langle f \rangle$ denotes the expectation value of the function f . For large N we get the asymptotic expression $\langle c \rangle \approx \frac{K\pi^2}{N^2}$. From the comparison with the optimal value $c_{\min} \approx \frac{\pi^2}{N^2}$ we note that entangling K systems and performing a joint measurement implies a reduction of the variance of a factor K in the estimation of the sum.

6. Conclusions

In this paper we addressed the estimation of an unknown quantum process that can possibly consist of a finite number of time steps. We formulated the search of the optimal quantum network for estimation as a semidefinite program, and used duality theory to give an alternative expression of the maximum payoff achieved by the optimal network. Using this result we proved a product rule for quantum metrology, showing that the individual strategies are sufficient to achieve the optimal joint estimate of a set of independent processes. In particular, the probability of success in the discrimination of K sets of processes is the product of the probabilities of success for each set.

It is easy to see that the product rule established here for joint estimation can also be generalized to the optimization of quantum networks for other tasks, such as the optimal cloning of independent sets of states. For example, in the case of cloning the product rule shows that the maximum fidelity for the joint cloning of K sets of states is the product of the maximum fidelities for each set, so that the optimal joint cloner is the product of the optimal individual cloners.

Acknowledgements. This work is supported the National Basic Research Program of China (973) 2011CBA00300 (2011CBA00302). The author gratefully acknowledges the hospitality of the Institute of Theoretical Computer Science and Communications, Chinese University of Hong Kong, where this work has been completed. A particular thanks goes to G Gutoski for pointing out the proof of strong duality in Ref. [35].

Appendix

Proof of theorem 1. Define the block diagonal matrices $T := \left(\bigoplus_{n=1}^N \Xi^{(n)} \right) \oplus \left(\bigoplus_{x \in \mathbf{X}} T_x^{(N)} \right)$ and $G = \left(\bigoplus_{n=1}^N 0_n \right) \oplus \left(\bigoplus_{x \in \mathbf{X}} G_x^{(N)} \right)$, where 0_i denotes the zero matrix in the i -th block. With these definitions, the optimization problem in Eq. (9) can be written as a semidefinite program in the standard form

$$\begin{aligned}
\gamma_{\max} &= \max_T && \text{Tr}[TG] \\
&\text{subject to} && T \geq 0 \\
&&& \mathcal{L}(T) = K
\end{aligned} \tag{16}$$

where \mathcal{L} is the Hermitian-preserving linear map defined by $\mathcal{L}(T) = \bigoplus_{j=0}^N R^{(j)}$ with

$$\begin{aligned} R^{(0)} &= \text{Tr}_{in,s_1}[\Xi^{(1)}] \\ R^{(1)} &= \text{Tr}_{in,s_2}[\Xi^{(2)}] - I_{out,s_1} \otimes \Xi^{(1)} \\ &\vdots \\ R^{(N-1)} &= \text{Tr}_{in,s_N}[\Xi^{(N)}] - I_{out,s_{N-1}} \otimes \Xi^{(N-1)} \\ R^{(N)} &= \left(\sum_{x \in \mathbf{X}} T_x \right) - I_{out,s_N} \otimes \Xi^{(N)}, \end{aligned}$$

and K is the block diagonal operator $K := \bigoplus_{j=0}^N K^{(j)}$ defined by $K^{(0)} = 1$ and $K^{(j)} = 0_j$ for every $j = 1, \dots, N$.

Using the duality of semidefinite programming we obtain

$$\begin{aligned} \gamma_{\max} \leq \gamma^* &:= \min_S \quad \text{Tr}[SK] \\ &\text{subject to } \mathcal{L}^\dagger(S) \geq G, \end{aligned} \tag{17}$$

where $S = \bigoplus_{j=0}^N S^{(j)}$ and \mathcal{L}^\dagger is the dual map defined by $\langle S, \mathcal{L}(T) \rangle = \langle \mathcal{L}^\dagger(S), T \rangle$ with $\langle S, T \rangle := \text{Tr}[S^\dagger T]$ is the Hilbert-Schmidt product. Using the definition of \mathcal{L}^\dagger , it is easy to check that $\mathcal{L}^\dagger(S) = \left(\bigoplus_{n=1}^N M_n \right) \oplus \left(\bigoplus_{x \in \mathbf{X}} M_x \right)$ where

$$\begin{aligned} M_1 &= I_{in,s_1} S^{(0)} - \text{Tr}_{out,s_1}[S^{(1)}] \\ M_2 &= I_{in,s_2} \otimes S^{(1)} - \text{Tr}_{out,s_2}[S^{(2)}] \\ &\vdots \\ M_N &= I_{in,s_N} \otimes S^{(N-1)} - \text{Tr}_{out,s_N}[S^{(N)}] \\ M_x &= S^{(N)} \quad \forall x \in \mathbf{X} \end{aligned}$$

Recalling the definition of K and G , the expression for γ^* becomes

$$\begin{aligned} \gamma^* &= \min_S \quad S^{(0)} \\ &\text{subject to } \begin{aligned} I_{in,s_1} S^{(0)} &\geq \text{Tr}_{out,s_1}[S^{(1)}] \\ I_{in,s_2} \otimes S^{(1)} &\geq \text{Tr}_{out,s_2}[S^{(2)}] \\ &\vdots \\ I_{in,s_N} \otimes S^{(N-1)} &\geq \text{Tr}_{out,s_N}[S^{(N)}] \\ S^{(N)} &\geq G_x^{(N)} \quad \forall x \in \mathbf{X}. \end{aligned} \end{aligned} \tag{18}$$

Note that $S^{(N)}$ must be positive, since we have $S^{(N)} \geq G_x^{(N)} \geq 0$. Consequently, $S^{(j)}$ must be positive for every $j = 0, \dots, N$. Moreover, there exists at least an operator S such that $\mathcal{L}^\dagger(S) > G$. For example, one can choose

$$\begin{aligned} S^{(N)} &= g_{\max} \prod_{n=1}^N (I_{out,s_n} \otimes I_{in,s_n}) \quad g_{\max} := \max_{\hat{x}, x \in \mathbf{X}} g(\hat{x}, x) \\ S^{(N-1)} &= 2 \text{Tr}_{out,s_N} \text{Tr}_{in,s_N}[S^{(N)}] \\ &\vdots \\ S^{(0)} &= 2 \text{Tr}_{out,s_1} \text{Tr}_{in,s_1}[S^{(s_1)}]. \end{aligned}$$

The existence of an operator S such that $\mathcal{L}^\dagger(S) > G$, along with the fact that the maximum payoff γ_{\max} is bounded by g_{\max} , implies that the hypotheses of Slater's theorem (see e.g. [35, 37]) on strong duality are satisfied. Hence, the optimum values for the primal and dual optimization problem coincide: $\gamma_{\max} = \gamma^*$.

Now, we show that the first N inequalities can be chosen to be equalities without loss of generality: we show that for every operator S satisfying the constraints there exists another operator \tilde{S} that achieves the equality in the first N constraints and has the same value of the objective function as S . To prove this statement, we proceed by induction. First, we define the operator $\tilde{S} := \sum_{j=0}^N \tilde{S}^{(j)}$ through the relations

$$\begin{aligned}\tilde{S}^{(0)} &:= S^{(0)} \\ \delta^{(1)} &:= I_{in,s_1} \tilde{S}^{(0)} - \text{Tr}_{out,s_1}[S^{(1)}] \geq 0 \\ \tilde{S}^{(1)} &:= S^{(1)} + \rho_1 \otimes \delta^{(1)}, \\ \tilde{S}^{(j)} &:= S^{(j)} \quad \forall j = 2, \dots, N\end{aligned}$$

where ρ_1 is an arbitrary quantum state in $\text{St}(\mathcal{H}_{out,s_1})$. Clearly, with this definition we have $\text{Tr}_{out,s_1}[\tilde{S}^{(1)}] = I_{in,s_1} \tilde{S}^{(0)}$, that is, \tilde{S} achieves the equality in the first constraint. Moreover, since $\delta^{(0)}$ is positive we have $I_{in,s_2} \otimes \tilde{S}^{(1)} \geq I_{in,s_2} \otimes S^{(1)} \geq \text{Tr}_{out,s_2}[S^{(2)}] \equiv \text{Tr}_{out,s_2}[\tilde{S}^{(2)}]$, namely \tilde{S} satisfies the second constraint. Hence, the operator \tilde{S} has the same objective value of S , satisfies all the constraints and achieves the equality in the first. Now, suppose that S achieves the equality in the first $k \geq 1$ constraints and define

$$\begin{aligned}\tilde{S}^{(j)} &:= S^{(j)} \quad \forall j = 1, \dots, k \\ \delta^{(k+1)} &:= I_{in,s_{k+1}} \tilde{S}^{(k)} - \text{Tr}_{out,s_{k+1}}[S^{(k+1)}] \geq 0 \\ \tilde{S}^{(k+1)} &:= S^{(k+1)} + \rho_{k+1} \otimes \delta^{(k+1)}, \\ \tilde{S}^{(j)} &:= S^{(j)} \quad \forall j = k+2, \dots, N\end{aligned}$$

where ρ_{k+1} is an arbitrary quantum state in $\text{St}(\mathcal{H}_{out,s_{k+1}})$. With this definition it is immediate to see that \tilde{S} has the same objective value of S , satisfies all constraints and achieves the equality in the first $k+1$ ones. By induction, we conclude that for every operator S satisfying the constraints there exists another operator \tilde{S} which achieves the equality in the first N constraints and has the same objective value. Defining $\lambda := \tilde{S}^{(0)}$ and $R := \tilde{S}^{(N)}/\lambda$ we then obtain the thesis of the theorem. ■

References

- [1] Caves C M 1981 *Phys. Rev. D* **23** 1693.
- [2] Wineland D J , Bollinger J J, Itano W M and Moore F L (1992) *Phys. Rev. A* **46R** 6797.
- [3] Derka R, Bužek V and Ekert A 1998 *Phys. Rev. Lett.* **80**, 1571.
- [4] Bužek V, Derka R and Massar S, 1999 *Phys. Rev. Lett.* **82** 2207.
- [5] Bagan E, Baig M, Brey A and Muñoz-Tapia R 2000 *Phys. Rev. Lett.* **85** 5230.
- [6] Peres A and Scudo P 2001 *Phys. Rev. Lett.* **86** 4160.
- [7] Chiribella G, D'Ariano G M, Perinotti P and Sacchi M F 2004 *Phys. Rev. Lett.* **93** 180503.
- [8] Bagan E, Baig M, and Muñoz-Tapia R 2004 *Phys. Rev. A* **70** 030301.
- [9] Hayashi M 2006 *Phys. Lett. A* **354** 183.

- [10] Giovannetti V, Lloyd S and Maccone L 2004 *Science* **306** 1330.
- [11] Giovannetti V, Lloyd S, and Maccone L 2006 *Phys. Rev. Lett.* **96** 010401.
- [12] van Dam W, D'Ariano G M, Ekert A, Macchiavello C and Mosca M 2007 *Phys. Rev. Lett.* **98** 090501.
- [13] Giovannetti V, Lloyd S, and Maccone L 2011 *Nature Photonics* **5** 222.
- [14] DiVincenzo D P, Leung D W and Terhal B M 2002 *IEEE Trans. Inf Theory* **48** 580.
- [15] Fung C-H and Chau H F 2008 *Phys. Rev. A* **78** 062308.
- [16] Helstrom C W 1976 *Quantum detection and estimation theory* (Academic Press, New York).
- [17] Holevo A S 1982 *Probabilistic and statistical aspects of quantum theory* (North-Holland, Amsterdam).
- [18] Feige U and Lovász L 1992 *Proceedings of the 24th ACM Symposium on the Theory of Computing* 733.
- [19] Raz R 1998 *SIAM Journal on Computing* **27**(3) 763.
- [20] Holenstein T 2007 *Proceedings of the 39th ACM Symposium on the Theory of Computing* 411.
- [21] Cleve R, Slofstra W, Unger F and Upadhyay S 2007 *Proceedings of the 22nd IEEE Conference on Computational Complexity, IEEE*.
- [22] Lee T, Shraibman A, and Spalek R 2008 *Proceedings of the 23rd IEEE Conference on Computational Complexity, IEEE*.
- [23] Mittal R and Szegedy M 2007 *Proceedings of FCT 2007, Lecture Notes in Computer Science* **4639** 435.
- [24] Lee T and Mittal R 2008 *Proceeding of ICALP '08, Lecture Notes in Computer Science* **5125** 674.
- [25] Chiribella G, D'Ariano G M and Perinotti P 2008 *Phys. Rev. Lett.* **101** 060401.
- [26] Chiribella G, D'Ariano G M, and Perinotti P 2009 *Phys. Rev. A* **80** 022339.
- [27] Gutoski G and Watrous J 2007 *Proceedings of STOC* **39** 565.
- [28] Yuen H P, Kennedy R S and Lax M 1975 *IEEE Trans. Inform. Theory* **IT-21** 125.
- [29] Bisio A, Chiribella G, D'Ariano G M and Perinotti P 2011 *Phys. Rev. A* **83** 022325.
- [30] Ježek M, Řeháček J and Fiurášek J 2002 *Phys. Rev. A* **65** 060301(R).
- [31] Hayashi M 2006 *Quantum Information: an Introduction* (Springer, Berlin).
- [32] Chiribella G, D'Ariano G M and Perinotti P 2009 *Proceedings of QCMC-08* 47.
- [33] Audenaert K M R, Calsamiglia J, Masanes Ll, Muñoz-Tapia R, Acín A, Bagan E and Verstraete F 2007 *Phys. Rev. Lett.* **98** 16050.
- [34] Chiribella G, D'Ariano G M and Perinotti P 2008 *Phys. Rev. Lett.* **101** 180501.
- [35] Gutoski G 2012 *J. Math. Phys.* **53** 032202.
- [36] Chiribella G, D'Ariano G M and M. F. Sacchi 2005 *Phys. Rev. A* **72** 042338.
- [37] Molina A and Watrous J 2011 *arXiv:1104.1140*.